

Assignment 10

Find your favorite news source and grab the article text.

In [1]:

```
# Excerpts taken from the Times's Interview with Trump
```

```
text = """TRUMP: Hi fellas, how you doing?
```

```
BAKER: Good. Good. How was your lunch [with Republican senators]?
```

```
TRUMP: It was good. We are very close. It's a tough — you know, health care. Look, I
```

```
But I think we are going to do O.K. I think we are going to see. I mean, one of my c
```

```
I mean, you think of Hillary Clinton, and you look, she went eight years — very capab
```

```
BAKER: March, March 2010.
```

```
TRUMP: So he was there for more than a year.
```

```
HABERMAN: Fourteen months.
```

```
TRUMP: And I'm here less than six months, so, ah, you know. Something to think about
```

```
BAKER: We wrote the same stories, though, in August of 2009. "Obama can't get it."
```

```
SCHMIDT: It died several times.
```

```
HABERMAN: Several times.
```

```
TRUMP: Well, it was a tough one. That was a very tough one.
```

```
BAKER: He lost that election [the 2010 midterms].
```

```
TRUMP: Nothing changes. Nothing changes. Once you get something for pre-existing con
```

```
HABERMAN: That's been the thing for four years. When you win an entitlement, you can
```

```
TRUMP: But what it does, Maggie, it means it gets tougher and tougher. As they get s
```

```
HABERMAN: Am I wrong in thinking — I've talked to you a bunch of times about this ov
```

```
TRUMP: Yes, yes. [garbled]
```

```
TRUMP: So I told them today, I don't want to do that. I want to either get it done c
```

BAKER: Did the senators want to try again?

TRUMP: I think so. We had a great meeting. Was I late?

[crosstalk]

TRUMP: It was a great meeting. We had 51 show up, other than John.

BAKER: Senator McCain.

TRUMP: That's a lot. Normally when they call for a meeting, you have like 20.

HABERMAN: How about the last one in June? Do you guys remember how many came?

TRUMP: Ah, 49. It was actually 48, but John McCain was there. But I guess we had 51.

BAKER: Who is the key guy?

TRUMP: Well, they are all key. The problem is we have 52 votes. Don't forget, you lo

HABERMAN: Where does it go from here, do you think?

TRUMP: Well, I say, let's not vote on repeal. Let's just vote on this. So first, the

HABERMAN: Next week.

TRUMP: Or Monday? Monday. And then they'll vote on this, and we'll see. We have some

And we have a very good group of people, and I think they want to get there. So we'

SCHMIDT: How's [Mitch] McConnell to work with?

TRUMP: I like him. I mean, he's good. He's good. It's been a tough process for him.

HABERMAN: He's taken on some water.

TRUMP: Yeah. It's been a tough process for him. This health care is a tough deal. I

This is a very tough time for them, in a sense, because of the importance. And I be

BAKER: Pretty tight.

TRUMP: I hope we don't have any grandstanders. I don't think we do.

[garbled]

TRUMP: I think it will be pretty bad for them if they did. I don't think we have any

HABERMAN: [In Paris], I don't think I've seen you look like you were enjoying yourse

TRUMP: I have had the best reviews on foreign land. So I go to Poland and make a spe

And then, went to France the following week, because it was the 100th year. [inaudible]

After that, it was fairly surprising. He [President Emmanuel Macron of France] called

HABERMAN: He was very deferential to you. Very.

TRUMP: He's a great guy. Smart. Strong. Loves holding my hand.

HABERMAN: I've noticed.

TRUMP: People don't realize he loves holding my hand. And that's good, as far as that

TRUMP: I mean, really. He's a very good person. And a tough guy, but look, he has to

[crosstalk]

TRUMP: At that note, the cameras are gone. I was standing there with him, with proba

HABERMAN: It was a very crowded [garbled].

TRUMP: And it was one of the most beautiful parades I have ever seen. And in fact, v

HABERMAN: I wondered if you were going to say that.

TRUMP: I've always thought of that.

HABERMAN: Really?

TRUMP: I've always thought of that. I've thought of it long before.

TRUMP: But the Bastille Day parade was — now that was a super-duper — O.K. I mean, t

These things are going on all day. It was a two-hour parade. They had so many differ

HABERMAN: It was beautiful.

TRUMP: And you are looking at the Arc [de Triomphe]. So we are standing in the most

People were surprised because I'd just come back from Hamburg. So I was back for the

TRUMP: She actually called me, and she said, um, "You know, I think we get along ver

TRUMP: It's been a long time. Nothing changes. Wait till you see what we're going to

HABERMAN: Sounds like it's going to be very interesting.

TRUMP: Much more interesting than anybody would understand.

HABERMAN: O.K.

BAKER: Will you go to Britain? Are you going to make a state visit to Britain? Are y

TRUMP: As to Britain?

BAKER: Yeah.

HABERMAN: Will you go there?

[crosstalk]

TRUMP: Ah, they've asked me. What was interesting – so, when Macron asked, I said: ‘

TRUMP: We had dinner at the Eiffel Tower, and the bottom of the Eiffel Tower looked

[crosstalk/garbled]

HABERMAN: You must have been so tired at, by that point.

TRUMP: Yeah. It was beautiful. We toured the museum, we went to Napoleon's tomb ...

[crosstalk]

TRUMP: Well, Napoleon finished a little bit bad. But I asked that. So I asked the p

[crosstalk/unintelligible]

TRUMP: Same thing happened to Hitler. Not for that reason, though. Hitler wanted to

[crosstalk]

But the Russians have great fighters in the cold. They use the cold to their advanta

So, we're having a good time. The economy is doing great.

SCHMIDT: The markets are doing great.

TRUMP: They're going to really go up if we do what we're doing. I mean, cut regulati

[crosstalk]

But I'm talking about for my time. I heard that Harry Truman was first, and then we

The energy stuff is going really well. We're going to be an exporter – we already a

BAKER: As long as we're on the record, a lot of people are curious about your conver

TRUMP: So, that dinner was a very long time planned dinner. And what it was was an e

TRUMP: So when we got there, it was with spouses, and when we got there, there were

BAKER: No, it was Julie [Hirschfeld Davis] and Glenn Thrush.

TRUMP: So, it was tremendous media. And we took a picture of everybody, the wives and

So, I was seated next to the wife of Prime Minister Abe [Shinzo Abe of Japan], who I

HABERMAN: Like, nothing, right? Like zero?

TRUMP: Like, not "hello."

HABERMAN: That must make for an awkward seating.

TRUMP: Well, it's hard, because you know, you're sitting there for—

HABERMAN: Hours.

TRUMP: So the dinner was probably an hour and 45 minutes.

TRUMP: You had an opera, and then you had a cocktail party for the people at the ope

It was like, that's where we're going. And I think it even said on the list, at the

O.K., so we're sitting at this massive table. And the wives are separated from their

And now Melania was sitting on the other side of the table, way down on the other en

HABERMAN: She had been the whole time?

TRUMP: Yes. She was sitting next to Putin.

BAKER: Does she speak Russian at all?

TRUMP: No. She speaks other languages.

TRUMP: She was sitting next to Putin and somebody else, and that's the way it is. So

HABERMAN: You did?

TRUMP: We talked about Russian adoption. Yeah. I always found that interesting. Bec

HABERMAN: The senators downstairs?

TRUMP: A lot of them. They said, "Who wouldn't have taken a meeting like that?"

BAKER: You asked them about it at lunch?

TRUMP: Nah, a couple of them. They — now, that was before Russia was hot, don't forget.

HABERMAN: Then can I ask you—

BAKER: Sorry to interrupt. The email, though, said something I thought was really interesting.

TRUMP: Well, I never saw the email. I never saw the email until, you know—

BAKER: Right, but now you have. So, what do you interpret that to mean, now that you have it?

TRUMP: Well, Hillary did the reset. Somebody was saying today, and then I read, when she was in the White House.

TRUMP: She was opposing sanctions. She was totally opposed to any sanctions for Russia.

BAKER: When was that?

HABERMAN: Do you remember when that was? I don't remember that.

TRUMP: I just saw it. I just saw it. She was opposed to sanctions, strongly opposed to sanctions.

HABERMAN: This is post-Crimea, I'm assuming? Is that what we would be talking about?

TRUMP: I don't really know. ... But in that time. And don't forget, Crimea was given a lot of money.

HABERMAN: It was sarin as well?

TRUMP: Sarin. And, and tremendous numbers of people were killed, young people, children.

HABERMAN: Yes.

TRUMP: You hate it, it's tough. Obama — you know, I can understand it in a way, but I don't like it.

BAKER: I do want to come out, on the email, now that you have seen that email that says that.

TRUMP: Well, I thought originally it might have had to do something with the payment to Russia.

HABERMAN: Which, which one?

SCHMIDT: The dossier.

TRUMP: The dossier.

HABERMAN: The dossier. Oh, yes.

TRUMP: Now, that was totally made-up stuff, and in fact, that guy's being sued by someone.

HABERMAN: Oh, I didn't know that.

TRUMP: I had a group of bodyguards, including Keith [Schiller] —

HABERMAN: Keith was there, right?

TRUMP: Keith was there. He said, "What kind of crap is this?" I went there for one c

TRUMP: When he [James B. Comey] brought it [the dossier] to me, I said this is real

HABERMAN: You said that to him?

TRUMP: Yeah, don't forget—

TRUMP: I said, this is — honestly, it was so wrong, and they didn't know I was just

SCHMIDT: Why do you think — why do you think he shared it?

TRUMP: I think he shared it so that I would — because the other three people left, a

TRUMP: So anyway, in my opinion, he shared it so that I would think he had it out th

SCHMIDT: As leverage?

TRUMP: Yeah, I think so. In retrospect. In retrospect. You know, when he wrote me th

[crosstalk]

BAKER: Do you think in hindsight, because of what's happened since then—

TRUMP: Comey wrote a letter.

HABERMAN: Which letter?

SCHMIDT: To you? To the F.B.I. staff or to you?

TRUMP: I thought it was to me, right?

BAKER: I think he wrote it to the staff, saying—

TRUMP: It might have been—

BAKER: That "the president has every right to fire me."

TRUMP: It might have been. It was just a very strange letter to say that.

BAKER: But do you think in hindsight, given that—

TRUMP: What was the purpose in repeating that?

BAKER: Do you think what's given that—

TRUMP: Do you understand what I mean? Why would somebody say, "He has every right to
[crosstalk]

TRUMP: It was very — a lot of people have commented that.

BAKER: Given what's happened since then, though, was it a political mistake to have

TRUMP: I think I did a great thing for the American people.

SCHMIDT: But look at the headache it's caused, you know?

TRUMP: It's okay. I have headaches, that's what I have, I have headaches. ... But you

HABERMAN: Do you wish you had done it on Day 1? When you got in? Because I honestly

TRUMP: Well, it could've been. It could've been. I feel like it was very dishonest v

SCHMIDT: What do you understand to be the four corners of what Mueller [Robert S. Mu

TRUMP: I don't know. Nobody has contacted me about anything.

TRUMP: Because I have done nothing wrong. A special counsel should never have been a

BAKER: Can we put that on the record?

TRUMP: Because so far, the only — yeah, you can put it down.

SCHMIDT: Was that [Attorney General Jeff] Sessions's mistake or [Deputy Attorney Gen

TRUMP: Look, Sessions gets the job. Right after he gets the job, he recuses himself.

BAKER: Was that a mistake?

TRUMP: Well, Sessions should have never recused himself, and if he was going to recu

HABERMAN: He gave you no heads up at all, in any sense?

TRUMP: Zero. So Jeff Sessions takes the job, gets into the job, recuses himself. I t

HABERMAN: Rosenstein.

TRUMP: Who is he? And Jeff hardly knew. He's from Baltimore.

TRUMP: Yeah, what Jeff Sessions did was he recused himself right after, right after

HABERMAN: I did, actually.

TRUMP: He was sitting in that chair. We had a wonderful meeting.

HABERMAN: Day before, right?

SCHMIDT: Did he want the job?

TRUMP: The day before! Of course, he was up here, and he wanted the job.

HABERMAN: And he made that clear to you? He would have—

TRUMP: So, now what happens is, he leaves the office. Rosenstein leaves the office.

HABERMAN: You mean at the hearing?

TRUMP: Yeah, he gave some answers that were simple questions and should have been s

HABERMAN: Why do you think it was? What do you think it was?

TRUMP: I don't know.

BAKER: What would cause you — what would be the line beyond which if Mueller went, y

TRUMP: Look, there are so many conflicts that everybody has. Then Rosenstein becomes

TRUMP: So think of this. Mike. He illegally leaks, and everyone thinks it is illegal

TRUMP: He said I said "hope" — "I hope you can treat Flynn good" or something like t

But even if he did — like I said at the news conference on the, you know, Rose Garde

TRUMP: And nothing was changed other than Richard Nixon came along. And when Nixon c

HABERMAN: Chris Wray.

TRUMP: He's highly thought of by everybody. I think I did the country a great service

BAKER: Did you shoo other people out of the room when you talked to Comey?

TRUMP: No, no.

BAKER: That time [inaudible] [Michael T.] Flynn —

TRUMP: No. That was the other thing. I told people to get out of the room. Why would

SCHMIDT: Did you actually have a one-on-one with Comey then?

TRUMP: Not much. Not even that I remember. He was sitting, and I don't remember even

ARABELLA KUSHNER: [enters room] Hi, Grandpa.

TRUMP: My granddaughter Arabella, who speaks — say hello to them in Chinese.

KUSHNER: Ni hao.

[laughter]

TRUMP: This is Ivanka. You know Ivanka.

IVANKA TRUMP: [from doorway] Hi, how are you? See you later, just wanted to come say

TRUMP: She's great. She speaks fluent Chinese. She's amazing.

BAKER: That's very impressive.

TRUMP: She spoke with President Xi [Jinping of China]. Honey? Can you say a few words

KUSHNER: Wo ai ni, Grandpa.

BAKER: That's great.

TRUMP: She's unbelievable, huh?

[crosstalk]

TRUMP: Good, smart genes.

[laughter]

TRUMP: So the bottom line is this. The country's doing well. We are, we are moving f

HABERMAN: It's a broad —

TRUMP: They're looking at a big picture.

BAKER: This is why I want to come back to that email, because, like — does it conce

TRUMP: By the way, it's everybody.

BAKER: Right, your point is that Democrats are trying to use this as an excuse, fine

TRUMP: You know, Peter, I didn't look into it very closely, to be honest with you.

BAKER: O.K.

TRUMP: I just heard there was an email requesting a meeting or something — yeah, rec

SCHMIDT: Did you know at the time that they had the meeting?

TRUMP: No, I didn't know anything about the meeting.

SCHMIDT: But you didn't—

TRUMP: It must have been a very important — must have been a very unimportant meetin

HABERMAN: No one told you a word, nothing? I know we talked about this on the plane

TRUMP: No, nobody told me. I didn't know noth— It's a very unimportant — sounded l

BAKER: But on the date you clinched the nominations with New Jersey and California a

TRUMP: Number one, remember, I made many of those speeches.

BAKER: People wondered about the timing.

TRUMP: Many of those speeches. I'd go after her all the time.

BAKER: Yeah, I know, but—

TRUMP: But there was something about the book, "Clinton Cash," came out.

BAKER: Yeah, a year earlier, though. But you were talking about—

TRUMP: But we were developing a whole thing. There was something about "Clinton Cash

TRUMP: Peter, that's all I did, was make those speeches about her. ... I don't think I

HABERMAN: [laughs] I'm sorry.

TRUMP: I mean, I was talking about, she deleted and bleached, which nobody does beca

HABERMAN: On Fifth Avenue—

TRUMP: I mean, look at what we have now. We have a director of the F.B.I., acting, v

HABERMAN: This is [Andrew] McCabe's wife, you mean?

TRUMP: McCabe's wife. She got \$700,000, and he's at the F.B.I. I mean, how do you th

BAKER: Should she be prosecuted now?

TRUMP: What?

BAKER: Should she be prosecuted now? Why, then, should she not be prosecuted now—

TRUMP: I don't want to say that. I mean, I don't want to say.

SCHMIDT: Last thing.

TRUMP: You understand what I mean, Peter.

BAKER: I know.

TRUMP: I mean, supposing they were able to give me additional — it wouldn't have helped.

SCHMIDT: Last thing, if Mueller—

TRUMP: And I couldn't have been better than the stuff I had. Obviously, because I was better.

SCHMIDT: Last thing, if Mueller was looking at your finances and your family finances—

HABERMAN: Would that be a breach of what his actual charge is?

TRUMP: I would say yeah. I would say yes. By the way, I would say, I don't — I don't think so.

SCHMIDT: But if he was outside that lane, would that mean he'd have to go?

[crosstalk]

HABERMAN: Would you consider—

TRUMP: No, I think that's a violation. Look, this is about Russia. So I think if he was outside that lane, would that mean he'd have to go?

HABERMAN: Would you fire Mueller if he went outside of certain parameters of what his charge is?

SCHMIDT: What would you do?

[crosstalk]

TRUMP: I can't, I can't answer that question because I don't think it's going to happen.

In [2]:

```
text
```

Out[2]:

```
'TRUMP: Hi fellas, how you doing?\n\nBAKER: Good. Good. How was your lunch [with Republican senators]?\n\nTRUMP: It was good. We are very close. It's a tough – you know, health care. Look, Hillary Clinton worked eight years in the White House with her husband as president and having majorities and couldn't get it done. Smart people, tough people – couldn't get it done. Obama worked so hard. They had 60 in the Senate. They had big majorities and had the White House. I mean, ended up giving away the state of Nebraska. They owned the state of Nebraska. Right. Gave it away. Their best senator did one of the greatest deals in the history of politics. What happened to him?\n\nBut I think we are going to do O.K. I think we are going to see. I mean, one of my ideas was repeal. But I certainly rather would get repeal and replace, because the next last thing I want to do is start working tomorrow morning on replace. And it is time. It is tough. It's a very narrow path, winding this way. You think you have it, and then you lose four on the other side because you gave. It is a brutal process. And it was for Democrats, in all fairness.\n\nI mean, you think of Hillary Clinton, and you look, she went eight years – very capable – went eight years as the fi
```

In [3]:

```
text = text.replace('\n\n','')
```

In [4]:

```
# Removing repetitive names identifying speaker
```

```
text = text.replace('TRUMP:', '')
text = text.replace('BAKER:', '')
text = text.replace('SCHMIDT:', '')
text = text.replace('HABERMAN:', '')
text
```

Out[4]:

```
' Hi fellas, how you doing? Good. Good. How was your lunch [with Republican senators]? It was good. We are very close. It's a tough – you know, health care. Look, Hillary Clinton worked eight years in the White House with her husband as president and having majorities and couldn't get it done. Smart people, tough people – couldn't get it done. Obama worked so hard. They had 60 in the Senate. They had big majorities and had the White House. I mean, ended up giving away the state of Nebraska. They owned the state of Nebraska. Right. Gave it away. Their best senator did one of the greatest deals in the history of politics. What happened to him? But I think we are going to do O.K. I think we are going to see. I mean, one of my ideas was repeal. But I certainly rather would get repeal and replace, because the next last thing I want to do is start working tomorrow morning on replace. And it is time. It is tough. It's a very narrow path, winding this way. You think you have it, and then you lose four on the other side because you gave. It is a brutal process. And it was for Democrats, in all fairness.\nI mean, you think of Hillary Clinton, and you look, she went eight years – very capable – went eight years as the first lady, and could not get health ca
```

1. Show the most common words in the article.

In [5]:

```
import spacy
from collections import Counter
import numpy as np
import pandas as pd
```

In [6]:

```
processor = spacy.load('en')
```

In [7]:

```
processed_text = processor(text)
processed_text
```

Out[7]:

Hi fellas, how you doing? Good. Good. How was your lunch [with Republican senators]? It was good. We are very close. It's a tough – you know, health care. Look, Hillary Clinton worked eight years in the White House with her husband as president and having majorities and couldn't get it done. Smart people, tough people – couldn't get it done. Obama worked so hard. They had 60 in the Senate. They had big majorities and had the White House. I mean, ended up giving away the state of Nebraska. They owned the state of Nebraska. Right. Gave it away. Their best senator did one of the greatest deals in the history of politics. What happened to him? But I think we are going to do O.K. I think we are going to see. I mean, one of my ideas was repeal. But I certainly rather would get repeal and replace, because the next last thing I want to do is start working tomorrow morning on replace. And it is time. It is tough. It's a very narrow path, winding this way. You think you have it, and then you lose four on the other side because you gave. It is a brutal process. And it was for Democrats, in all fairness. I mean, you think of Hillary Clinton, and you look, she went eight years – very capable – went eight years as the first lady, and could not

In [8]:

```
counts = Counter()

for token in processed_text:
    counts[token.text] += 1

list(counts.most_common()[ :20])
```

Out[8]:

```
[(',', 611),
 ('.', 600),
 ('I', 319),
 ('_', 253),
 ('the', 251),
 ('you', 169),
 ('a', 167),
 ('to', 161),
 ('was', 154),
 ('that', 147),
 ('it', 140),
 ('and', 118),
 (''s', 117),
 ('of', 111),
 ('?', 107),
 ('n't', 104),
 ('-', 87),
 ('do', 87),
 ('he', 85),
 ('have', 82)]
```


In [9]:

```
# Alternately

counts = Counter()

for sentence in processed_text.sents:
    for token in sentence:
        counts[token.text] += 1

# counts.most_common()[:20]
counts.most_common(20)
```

Out[9]:

```
[(',', 611),
 ('.', 600),
 ('I', 319),
 ('_', 253),
 ('the', 251),
 ('you', 169),
 ('a', 167),
 ('to', 161),
 ('was', 154),
 ('that', 147),
 ('it', 140),
 ('and', 118),
 (''s', 117),
 ('of', 111),
 ('?', 107),
 ('n't', 104),
 ('-', 87),
 ('do', 87),
 ('he', 85),
 ('have', 82)]
```

The wordcount for 'I' is almost twice that of 'you'.

2. Show the most common words under a part of speech. (i.e. NOUN: {'Bob':12, 'Alice':4,})

In [10]:

```
from collections import defaultdict
```

In [16]:

```
# counts = Counter()
# part_of_speech = {}
part_of_speech = defaultdict(Counter)

for sentence in processed_text.sents:
    for token in sentence:

        part_of_speech[token.pos_][token.text] += 1
#         print ({'token': token, 'pos': token.pos_})
```

In [17]:

```
for key, value in part_of_speech.items():
    print(key, value.most_common(10))
```

```
SPACE [(' ', 2), ('\n', 2)]
INTJ [('Well', 11), ('Yeah', 10), ('No', 10), ('like', 5), ('Hi', 4),
('yes', 4), ('no', 4), ('yeah', 4), ('Like', 4), ('hello', 4)]
NOUN [('people', 45), ('what', 38), ('thing', 27), ('time', 18), ('som
ething', 17), ('meeting', 16), ('lot', 16), ('things', 14), ('years',
13), ('way', 11)]
PUNCT [(',', 611), ('.', 600), ('_', 253), ('?', 107), ('-', 87), (']'
, 63), ('[', 53), ('"', 37), ('"', 37), ('-', 13)]
ADV [('n't', 102), ('very', 42), ('So', 42), ('there', 37), ('then', 2
6), ('so', 23), ('when', 20), ('just', 20), ('not', 19), ('really', 19
)]
PRON [('I', 319), ('you', 169), ('it', 140), ('he', 85), ('we', 63), (
'It', 51), ('they', 51), ('You', 32), ('me', 32), ('He', 29)]
VERB [('was', 154), ('do', 87), ('have', 82), ('know', 67), ('is', 61)
, ('s', 59), ('think', 59), ('did', 57), ('had', 52), ('said', 48)]
ADJ [('tough', 22), ('great', 19), ('other', 18), ('that', 18), ('good
', 17), ('my', 14), ('which', 11), ('right', 10), ('big', 9), ('many',
9)]
ADP [('of', 110), ('to', 61), ('in', 53), ('that', 52), ('with', 47),
('about', 45), ('for', 40), ('at', 37), ('on', 31), ('because', 28)]
```

Russia was the main topic. Interestingly, Clinton was mentioned twice as many times as Comey. She was also referred to by her first name half the time. Obama was only referred to by his last name.

3. Find a subject/object relationship through the dependency parser in any sentence.

In [18]:

```
text = text.replace('\n\n', '')
```

In [19]:

```
# Establish parsing function

def pr_tree(word, level):
    if word.is_punct:
        return
    for child in word.lefts:
        pr_tree(child, level+1)
    print('\t'* level + word.text + ' - ' + word.dep_)
    for child in word.rights:
        pr_tree(child, level+1)
```

In [20]:

```
for sentence in processed_text.sents:
    pr_tree(sentence.root, 0)
    print('-----')
```

```

-
Hi - ROOT
    fellas - npadvmod
        how - advmod
        you - nsubj
    doing - ccomp
-----
Good - ROOT
-----
Good - ROOT
-----
    How - nsubj
was - ROOT
    your - poss
    lunch - nsubj
        with - prep
                Republican - amod
                senators - pobj
-----
    to - nsubj
```

Sentence 1:

'Hi' is the root, followed by 'fellas' which is related to both adverbial modifier 'how' and subject 'you' and 'doing' is the clausal complement.

Sentences 2-3:

'Good' and 'Good' are both roots as they are sentences in and of themselves.

Sentence 4:

'was' is the root verb, 'How' and 'lunch' are both subjects, with 'your' as a possessive to 'lunch'. 'with' is identified as related to 'lunch' as a preposition, with 'senators' as object of the preposition and 'Republican' adjective.

4. Show the most common Entities and their types.

In [202]:

```
# for entity in processed_text.ents:  
#     print(entity, entity.label_)
```

In [203]:

```
# # Most common entity types  
  
# # counts = defaultdict()  
# counts = Counter()  
  
# for entity in processed_text.ents:  
#     counts[entity.label_] += 1  
  
# dict(counts)  
  
# # for key, value in counts.items():  
# #     print(key.most_common(20),value)
```

In [204]:

```
# # Most common entities  
  
# counts = Counter()  
  
# for entity in processed_text.ents:  
#     counts[entity.text] += 1  
  
# counts.most_common(10)
```

In [21]:

```
# Most common entities and their types  
  
counts = Counter()  
  
for entity in processed_text.ents:  
    counts[entity.text, entity.label_] += 1
```

In [22]:

```
counts.most_common()
```

Out[22]:

```
[(('Russia', 'GPE'), 29),  
 (('one', 'CARDINAL'), 17),  
 (('Hillary Clinton', 'PERSON'), 9),  
 (('France', 'GPE'), 6),  
 (('F.B.I.', 'ORG'), 6),  
 (('Trump', 'PERSON'), 5),  
 (('O.K.', 'GPE'), 5),  
 (('Putin', 'PERSON'), 5),  
 (('Mueller', 'PERSON'), 5),  
 (('eight years', 'DATE'), 4),  
 (('Obama', 'ORG'), 4),  
 (('Democrats', 'NORP'), 4),  
 (('today', 'DATE'), 4),  
 (('Paris', 'GPE'), 4),  
 (('Russian', 'NORP'), 4),  
 (('Comey', 'PERSON'), 4),  
 (('Jeff Sessions', 'PERSON'), 4),  
 (('Rosenstein', 'PERSON'), 4)].
```

Russia and Hillary Clinton are the most common entities and types. One may just be a commonly used word in speech.

5. Find Entites and their dependency (hint: entity.root.head)

In [23]:

```
for entity in processed_text.ents:
    pr_tree(entity.root.head, 0)
    print('-----')
```

```

    Republican - amod
senators - pobj
-----
    Look - parataxis
        Hillary - compound
    Clinton - nsubj
worked - ROOT
        eight - nummod
    years - dobj
    in - prep
            the - det
            White - compound
        House - pobj
    with - prep
            her - poss
        husband - pobj
            as - prep
                president - pobj
                    and - cc
                    husband - dobj
                    Clinton - nsubj
                    worked - ROOT
```

In [208]:

```
# for entity in processed_text.ents:
#     pr_tree(entity.root, 0)
#     print('-----')
```

6. Find the most similar nouns (chunks) in the article

In [24]:

```
states = processed_text.vocab['states']
government = processed_text.vocab['government']

states.similarity(government)
```

Out[24]:

0.5907375800423037

In [25]:

```
vote = processed_text.vocab['vote']  
  
government.similarity(vote)
```

Out[25]:

0.44072801707864545

In [280]:

```
# # find noun chunks  
# noun_chunks = []  
  
# for noun_chunk in processed_text.noun_chunks:  
#     noun_chunks.append(noun_chunk)  
  
# print(list(noun_chunks))
```

...

In [323]:

```
# # get vectors for noun_chunks  
  
# noun_chunks = []  
# for noun_chunk in processed_text.noun_chunks:  
#     noun_dict = {'noun_chunk': noun_chunk, 'noun_vector': noun_chunk.vector}  
# #     print(noun_dict)  
#     noun_chunks.append(noun_dict)  
  
# noun_chunks[:5]
```

...

In [26]:

```
# progress bar  
from tqdm import tqdm_notebook as tqdm  
  
# spacy  
import spacy  
processor = spacy.load('en')
```

In [348]:

```
# text = str.lower(text)  
# text
```

...

In [351]:

```
# processed_text = processor(text)
# processed_text
```

...

In [32]:

```
# most similar noun_chunks
most_similar = ''
# best similarity score
best_similarity = 0

# establish noun_1, noun_2 in processed_text.noun_chunks
for noun_1 in tqdm(processed_text.noun_chunks):
    for noun_2 in processed_text.noun_chunks:
        # skip comparing the same noun
        if noun_1.lower_ == noun_2.lower_:
            # if noun_1.lower_ == noun_2.lower_:
            # skip
            continue
        similar = noun_1.similarity(noun_2)
        # some similarities were greater than 1, which doesn't make sense
        if similar > 1:
            continue
        # show noun_chunks with best similarity score
        if similar > best_similarity:
            best_similarity = similar
            most_similar = (noun_1, noun_2)
            print(best_similarity, most_similar)
```

×

2196it [04:14, 7.71it/s]

```
0.677260983949 (you, your lunch)
0.716595202094 (you, It)
0.746166639379 (you, We)
0.755075812487 (you, What)
0.757896225621 (you, this way)
0.78929663874 (you, yourself)
0.808562911835 (you, your point)
0.889949134187 (you, You point)
0.919694455945 (It, as it)
0.999999928204 (It, saying— It)
0.999999992546 (the Senate, the senators)
```

The Senate and senators have the most similarity, and appropriately so. For some reason, some scores were greater than 1 which made no sense, so those were just skipped and not included in the model.

In []: